

MAT-8193US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Keiko Morii; Yoshio Ohno
Serial No.: To be assigned
Filed: Herewith
FOR: METHOD FOR SPEECH
RECOGNITION, APPARATUS FOR
THE SAME, AND VOICE
CONTROLLER

: Art Unit:
: Examiner:

PATENT

Jc997 U.S. PTO
09/975918
10/12/01

CLAIM TO RIGHT OF PRIORITY

Assistant Commissioner for Patents
Washington, D.C. 20231

S I R :

Pursuant to 35 U.S.C. 119, Applicant's claim to the benefit of filing of prior Japanese Patent Application No. 2000-311599, filed October 12, 2000, as stated in the inventor's Declaration, is hereby confirmed.

A certified copy of each of the above-referenced applications is enclosed.

Respectfully submitted,

Lawrence E. Ashery, Reg. No. 34,515
Attorney for Applicants

LEA/jam

Enclosures: Certified Copy of Japanese Application
Dated: October 12, 2001

Suite 301
One Westlakes, Berwyn
P.O. Box 980
Valley Forge, PA 19482-0980
(610) 407-0700

The Assistant Commissioner for Patents is hereby authorized to charge payment to Deposit Account No. 18-0350 of any fees associated with this communication.

EXPRESS MAIL Mailing Label Number: EL 741592806 US

Date of Deposit: October 12, 2001

I hereby certify that this paper and fee are being deposited, under 37 C.F.R. § 1.10 and with sufficient postage, using the "Express Mail Post Office to Addressee" service of the United States Postal Service on the date indicated above and that the deposit is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

Kristen Foley
Kristen Foley

日 本 国 特 許 庁
JAPAN PATENT OFFICE

JC997 U.S. PTO
09/975918
10/12/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年10月12日

出 願 番 号

Application Number:

特願2000-311599

出 願 人

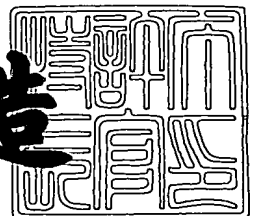
Applicant(s):

松下電器産業株式会社

2001年 7月27日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3067198

【書類名】 特許願

【整理番号】 2931020073

【提出日】 平成12年10月12日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/06

【発明者】

 【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技
研株式会社内

 【氏名】 森井 景子

【発明者】

 【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技
研株式会社内

 【氏名】 大野 剛男

【特許出願人】

 【識別番号】 000005821

 【氏名又は名称】 松下電器産業株式会社

【代理人】

 【識別番号】 100097445

 【弁理士】

 【氏名又は名称】 岩橋 文雄

【選任した代理人】

 【識別番号】 100103355

 【弁理士】

 【氏名又は名称】 坂口 智康

【選任した代理人】

 【識別番号】 100109667

 【弁理士】

 【氏名又は名称】 内藤 浩樹

【手数料の表示】

【予納台帳番号】 011305

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809938

【書類名】 明細書

【発明の名称】 音声認識方法及びその装置、ならびに音声制御装置

【特許請求の範囲】

【請求項 1】 学習パターンと入力音声と比較する音声認識方法において、学習話者の特徴で分類した複数の学習パターンを予め保持し、ユーザの第一の発声により前記複数の学習パターンから学習パターンの選択を行い、前記ユーザの第一の発声と前記選択された学習パターンからユーザの発声に対する音声スペクトル周波数の歪曲係数を求め、前記選択された学習パターンと前記歪曲係数を用いて、第一の発声に続く入力音声を認識することを特徴とする音声認識方法。

【請求項 2】 学習パターンの分類は、学習話者の年齢によることを特徴とする請求項 1 記載の音声認識方法。

【請求項 3】 学習パターンの分類は、地域によることを特徴とする請求項 1 記載の音声認識方法。

【請求項 4】 分類した複数の学習パターンは、特徴カテゴリ毎の発声の平均値、特徴カテゴリ毎の発声の共分散と、全学習話者の発声の共分散からなることを特徴とする請求項 1 から 3 のいずれかに記載の音声認識方法。

【請求項 5】 学習パターンの選択は、ユーザの第一の発声と、特徴カテゴリ毎の発声の平均値と全学習話者の発声の共分散からなる学習パターンとの距離計算により学習パターンの選択を行い、選択された特徴カテゴリの発声の平均値と特徴カテゴリの発声の共分散からなる学習パターンを出力することを特徴とする請求項 4 記載の音声認識方法。

【請求項 6】 学習パターンと入力音声と比較する音声認識装置において、入力された音声を音響分析して音響パラメータを求める音響分析部と、学習話者の特徴で分類した複数の学習パターンを予め保持する特徴別パターン格納部と、ユーザの第一の発声により前記複数の学習パターンから学習パターンの選択を行う特徴別パターン選択構成部と、前記ユーザの第一の発声の音響パラメータと前記選択された学習パターンを用いて前記音響パラメータに対する音声スペクトル周波数の歪曲係数を求める話者適応処理部と、前記歪曲係数と前記選択された学習パターンを用いて、第一の発声に続く入力音声を認識する音声認識部とを具備す

ることを特徴とする音声認識装置。

【請求項 7】 特徴別パターン選択構成部は、ユーザの第一の発声と、特徴カテゴリ毎の発声の平均値と全学習話者の発声の共分散からなる学習パターンとの距離計算により学習パターンの選択を行い、選択された特徴カテゴリの発声の平均値と特徴カテゴリの発声の共分散からなる学習パターンを出力することを特徴とする請求項 6 記載の音声認識装置。

【請求項 8】 請求項 6 または 7 記載の音声認識装置を用いたことを特徴とする音声制御装置。

【請求項 9】 発声によるパターン選択時にはあらかじめ用意されている 1 つもしくは複数の既知語のうちの 1 語を用い、当該既知語によって制御対象の機器選択を行うことを特徴とする請求項 8 記載の音声制御装置。

【請求項 10】 発声によるパターン選択時の既知語は、ユーザーが登録できることを特徴とする請求項 8 記載の音声制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、幅広い話者を対象とした音声認識方法及びその装置、ならびに音声制御装置にかかわる。

【0002】

【従来の技術】

認識話者を、年齢などの特徴によって、あらかじめ分類してパターンを学習する音声認識方法は、電子通信学会論文誌 '80/12 Vol.J63-D No.12 pp.1002-1009 や特開平 10-282986 号公報などに開示されている。

【0003】

また、認識話者に対して単一のパターンを用い、話者の音声のスペクトル周波数を歪曲させる話者適応方法も、Wakita, H. "Normalization of Vowels by Vocal-Tract Length and Its Application to Vowel Identification," IEEE Trans . ASSP 25(2): pp.183-192 (1977) などで、広く検討されている。

【0004】

一方、細かい話者の特徴を吸収することができる話者適応の方法としては、最大事後確率推定法（MAP 推定法）などが知られており、電子情報通信学会技術研究報告 Vol.93 No.427 pp.39-46（SP93-133, 1993）などに開示されているが、この方法では、適応化のための学習音声我非常に少ない場合、例えば一文章だけでは、適応化による改善を得ることは出来ないと明示されている。

【0005】

【発明が解決しようとする課題】

これらの従来技術を用いても、適応のための話者の労力が少ない、すなわちユーザーの1単語から数単語程度の数少ない発声を用いて、ユーザーの細かい特徴を吸収することは困難であり、得られる音声認識性能も不十分であった。

【0006】

本発明は、ユーザーの数少ない発声によって、細かいユーザーの特徴を吸収し、音声認識性能を向上することを目的とする。

【0007】

【課題を解決するための手段】

この課題を解決するための本発明は、入力された音声を音響分析して音響パラメータを求める音響分析部と、学習話者の特徴で分類した複数の学習パターンを予め保持する特徴別パターン格納部と、ユーザの第一の発声により前記複数の学習パターンから学習パターンの選択を行う特徴別パターン選択構成部と、前記ユーザの第一の発声の音響パラメータと前記選択された学習パターンを用いて前記音響パラメータに対する音声スペクトル周波数の歪曲係数を求める話者適応処理部と、前記歪曲係数と前記選択された学習パターンを用いて、第一の発声に続く入力音声を認識する音声認識部とを具備することによって、数少ない発声で簡便かつ適切な話者適応を行う音声認識方法である。

【0008】

【発明の実施の形態】

本発明の請求項1に記載の発明は、学習パターンと入力音声を比較する音声認識方法において、学習話者の特徴で分類した複数の学習パターンを予め保持し、ユーザの第一の発声により前記複数の学習パターンから学習パターンの選択を行

い、前記ユーザの第一の発声と前記選択された学習パターンからユーザの発声に対する音声スペクトル周波数の歪曲係数を求め、前記選択された学習パターンと前記歪曲係数を用いて、第一の発声に続く入力音声を認識することを特徴とするものであって、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0009】

請求項2に記載の発明は、請求項1記載の音声認識方法において、学習パターンの分類は、学習話者の年齢によることを特徴とするもので、特徴カテゴリを学習話者の年齢にすることにより、学習パターンの選択が確実に行うことができ、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0010】

請求項3に記載の発明は、請求項1記載の音声認識方法において、学習パターンの分類は、地域によることを特徴とするもので、特徴カテゴリを地域毎に分類するもので、話者の訛りや言語によって分類することにより、学習パターンの選択が確実に行うことができ、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0011】

請求項4に記載の発明は、請求項1から3のいずれかに記載の音声認識方法において、分類した複数の学習パターンは、特徴カテゴリ毎の発声の平均値、特徴カテゴリ毎の発声の共分散と、全学習話者の発声の共分散からなることを特徴とするもので、学習パターンの選択が確実に行うことができ、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0012】

請求項5に記載の発明は、請求項4記載の音声認識方法において、学習パターンの選択は、ユーザの第一の発声と、特徴カテゴリ毎の発声の平均値と全学習話者の発声の共分散からなる学習パターンとの距離計算により学習パターンの選択を行い、選択された特徴カテゴリの発声の平均値と特徴カテゴリの発声の共分散

からなる学習パターンを出力することを特徴とするもので、カテゴリの選択時と認識時の学習パターンの組み合わせを変えることにより、学習パターンの選択が確実に行うことができ、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0013】

本発明の請求項6に記載の発明は、学習パターンと入力音声と比較する音声認識装置において、入力された音声を音響分析して音響パラメータを求める音響分析部と、学習話者の特徴で分類した複数の学習パターンを予め保持する特徴別パターン格納部と、ユーザの第一の発声により前記複数の学習パターンから学習パターンの選択を行う特徴別パターン選択構成部と、前記ユーザの第一の発声の音響パラメータと前記選択された学習パターンを用いて前記音響パラメータに対する音声スペクトル周波数の歪曲係数を求める話者適応処理部と、前記歪曲係数と前記選択された学習パターンを用いて、第一の発声に続く入力音声を認識する音声認識部とを具備するもので、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0014】

請求項7に記載の発明は、請求項6記載の音声認識装置において、特徴別パターン選択構成部は、ユーザの第一の発声と、特徴カテゴリ毎の発声の平均値と全学習話者の発声の共分散からなる学習パターンとの距離計算により学習パターンの選択を行い、選択された特徴カテゴリの発声の平均値と特徴カテゴリの発声の共分散からなる学習パターンを出力することを特徴とするもので、学習パターンの選択が確実に行うことができ、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声認識性能を向上させるという作用を有する。

【0015】

請求項8に記載の発明は、請求項6または7記載の音声認識装置を用いたことを特徴とするもので、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声制御性能を向上させるという作用を有する。

【0016】

請求項9に記載の発明は、請求項8に記載の音声制御装置において、複数の学

習パターンの中から認識に使用するパターンを選択する際のユーザーの発声は、1つもしくは複数の既知語のうちの1語を用い、当該既知語によって制御対象の機器選択を行うことを特徴とするものであって、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声制御性能を向上させるという作用を有する。

【0017】

請求項10に記載の発明は、請求項8に記載の音声制御装置において、複数の学習パターンの中から認識に使用するパターンを選択する際のユーザーの発声は、ユーザーがあらかじめ登録しておいた1つもしくは複数の語のうちの1語であって、当該発声語によって制御対象の機器選択を行うことを特徴とするものであって、ユーザーの数少ない発声で簡便かつ適切な話者適応を行うことにより音声制御性能を向上させるという作用を有する。

【0018】

以下、本発明の実施の形態について、図を用いて説明する。

【0019】

(実施の形態1)

図1に、本発明の実施の形態1における音声制御装置のブロック構成図を示す。図1において、1はユーザーの音声を入力する音声入力部、2はユーザーの発声音声の音響分析を行う音響分析部、3はあらかじめ学習話者の特徴に応じて分類して学習されている特徴別パターンを格納する特徴別パターン格納部、4はユーザーの発声音声の音響パラメータによって特徴別パターンを選択、再構成する特徴別パターン選択構成部、5は特徴別パターン選択構成部4によって決定された特徴別パターンと音響分析部2から入力されたユーザーの発声音声の音響パラメータとを比較して、ユーザーの発声音声のスペクトル周波数の歪曲度を示す変換係数を決定する話者適応処理部、6は特徴別パターン選択部4で決定された特徴別パターンと話者適応処理部5で決定された変換係数を用いて認識処理を行う音声認識部、7は音声認識部6で得られた結果に応じた制御信号を出力する制御信号出力部である。

【0020】

ユーザーは、まずパターン選択、話者適応、及び機器選択のために既定された語を発声し、続けて認識対象語を発声する。本実施例においては、例えば、「テレビ音__大きく」であるとか、「照明__消して」と発声する。この発声内容には、パターン選択と話者適応のための既定された機器選択語と、それに続けて制御内容を示す1語が含まれる。

【0021】

音声入力部1からは、ユーザーの発声がマイクロフォンを経由してA/D変換されて、デジタル信号として音声制御装置に取り込まれる。音響分析部2においては音響パラメータとしてLPCケプストラム係数を求める。本実施例では、音響パラメータとしてLPCケプストラムを使用する例を挙げているが、MFCC (mel frequency cepstral coefficients)など他の音響パラメータでも同様の効果が得られる。

【0022】

特徴別パターン格納部3の詳細構成を図2に示す。特徴別パターン格納部3にあらかじめ格納されている特徴別パターンは、学習話者の年齢、学習話者が最も長期にわたって住んでいるもしくは住んでいた地域や、学習話者の母国語によって分類することが考えられるが、ここでは学習話者の年齢による場合を説明する。

特徴別パターン格納部3にあらかじめ格納されている特徴別パターンは、予め用意してある学習音声データを話者の年齢によって分類し、各特徴カテゴリの学習話者の発声の平均値、各特長カテゴリの学習話者の発声の共分散、全学習話者の発声の共分散、各データから構成される。

【0023】

本実施例では12歳以下、13歳～64歳、65歳以上の3つに分類し、それぞれ、12歳以下をカテゴリ番号201、13歳～64歳をカテゴリ番号202、65歳以上をカテゴリ番号203とする。101は全学習話者の発声の共分散である。

【0024】

また、111はカテゴリ番号201の学習話者の発声の平均値であり、112

はカテゴリ番号 2 0 2 の学習話者の発声の平均値、1 1 3 はカテゴリ番号 2 0 3 の学習話者の発声の平均値である。そして、1 2 1 はカテゴリ番号 2 0 1 の学習話者の発声の共分散であり、1 2 2 はカテゴリ番号 2 0 2 の学習話者の発声の共分散、1 2 3 はカテゴリ番号 2 0 3 の学習話者の発声の共分散である。

【 0 0 2 5 】

特徴別パターン選択構成部 4 においては、音響分析部 2 で得られた L P C ケプストラム係数のうち、第 1 語目の発声の部分について、予め用意されている特徴別パターンとの距離計算処理を行い、後続語の認識に使用すべき特徴別パターンの決定を行う。ここで、使用する距離尺度は、マハラノビス汎距離を基本としてこれを一次判別関数として展開したもの（簡易型マハラノビス距離と呼ぶ、例えば、特開昭 6 0 - 6 7 9 9 6 号公報に開示されている）を使用する。なお本実施例では、統計的距離尺度である簡易型マハラノビス距離を用いる例を記述しているが、ベイズ判定などの実施例とは異なる統計的距離尺度を使用しても良いし、隠れマルコフモデルの出力確率を使用しても同様の効果が得られる。

【 0 0 2 6 】

図 3 に特徴別パターン選択構成部 4 の処理のフローチャートを示す。S 0 1 において入力音声の音響パラメータの読み込みを行う。すなわちここでは入力音声の音響分析を行って得られる L P C ケプストラム係数を読み込む。S 0 2 においては、選択対象パターンの読み込みを行う。すなわちここでは図 2 に示す特徴別パターン格納部 4 に格納されているパターンのうち、カテゴリ番号 2 0 1 の学習話者の発声の平均値 1 1 1 と学習話者全員の発声の共分散 1 0 1 をカテゴリ番号 2 0 1 の選択用特徴別パターンとして、カテゴリ番号 2 0 2 の学習話者の発声の平均値 1 1 2 と学習話者全員の発声の共分散 1 0 1 をカテゴリ番号 2 0 2 の選択用特徴別パターンとして、そして、カテゴリ番号 2 0 3 の学習話者の発声の平均値 1 1 3 と学習話者全員の発声の共分散 1 0 1 をカテゴリ番号 2 0 3 の選択用特徴別パターンとして読み込む。

【 0 0 2 7 】

S 0 3 においては各特徴別パターンを用いて、特徴別パターン選択と機器選択を兼ねるための「テレビ」、「ビデオ」、「エアコン」、「照明」の 4 語を認識

対象語として距離計算を行って、特徴別パターン毎に、上記4語の認識対象語と、ユーザーの入力音声との距離を得る。本実施例では4語の認識対象語の例になっているが、認識対象語が10語程度でも、1語でも同様の効果が得られる。

【0028】

S04においては、S03で得られた距離を比較する。S05においては、S04で得られた比較結果より、最も距離が近いパターン番号と制御対象機器を示す語彙の認識結果を選定する。

【0029】

S06においては、S05において選定された最も距離が近いパターンに応じて被選択パターンの再構成を行う。すなわち、S04において選定された特徴別パターンが番号201であったならば、カテゴリ番号201の学習話者の発声の平均値111とカテゴリ番号201の学習話者の発声の共分散121を被選択特徴別パターンとして、また、選定された特徴別パターンが番号202であったならば、カテゴリ番号202の学習話者の発声の平均値112とカテゴリ番号202の学習話者の発声の共分散122を被選択特徴別パターンとして、そして、選定された特徴別パターンが番号203であったならば、カテゴリ番号203の学習話者の発声の平均値113とカテゴリ番号203の学習話者の発声の共分散123を被選択特徴別パターンとして再構成して、処理を終了する。

【0030】

話者適応処理部5では、先に音響分析部2で計算された入力音声のLPCケプストラム係数のうち、第1語目の発声の部分を用いて、Oppenheim, A.V. and Johnson, D.H. "Discrete Representation of Signals," Proc. IEEE 60 (6): 681-691(1972)に開示されている(数1)のオッペンハイムの方法によってスペクトル周波数を歪曲しながら、特徴別パターン選択構成部4で決定された特徴別パターンを用い、同じく特徴別パターン選択構成部4で決定された制御対象の認識結果語彙を辞書として、発声の距離尺度を計算し、最尤の結果を与える周波数移動の係数を(数2)に則って決定する。

【0031】

【数 1】

$$\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}$$

【0032】

【数 2】

$$\hat{\alpha} = \arg \max_{\alpha} P(X^{\alpha} | \alpha, \theta)$$

【0033】

図4に話者適応処理部5のフローチャートを示す。S11では演算対象のスペクトル周波数歪曲係数の初期値を3個設定する。S12ではS11で設定されたスペクトル周波数歪曲係数に応じて、音響分析部2から得られる、ユーザーの発声のLPCケプストラム係数のうちの第1語目の発声の部分に対して、スペクトル歪曲係数の演算を行い、スペクトルが歪曲された、3通りのLPCケプストラム係数を得る。S13においては、特徴パターン別選択構成部4において決定された被選択特徴別パターンと制御対象機器を示す語彙の認識結果を読み込む。

【0034】

S14においては、S12において得られた3通りのLPCケプストラム係数とS13で得られた被選択特徴別パターンとの距離計算を、S13で得られた制御対象機器を示す語彙の認識結果に対して行う。S16においては、得られた3つの距離尺度のうち、最尤すなわち最も近い距離尺度が得られるスペクトル周波数歪曲係数が3個の中心か否かを判定する。S16における判定で最尤距離を与えるスペクトル周波数歪曲係数が3個の中心でないならば、S18において、最尤距離を与えるスペクトル周波数歪曲係数が、3個の歪曲係数のうちの最大値か否かを判定する。最尤距離を与えるスペクトル周波数歪曲係数が3個の歪曲係数のうちの最大値である場合には、S19において演算対象歪曲係数3個全てに0.02を加算して、S12のスペクトル周波数歪曲演算に戻る。最尤距離を与えるスペクトル周波数歪曲係数が3個の歪曲係数のうちの最大値でない場合には、

S 2 0 において演算対象歪曲係数 3 個全てから 0. 0 2 を減算して、S 1 2 のスペクトル周波数歪曲演算に戻る。こうして繰り返し処理を行う間に、S 1 6 における判定で、最尤距離を与えるスペクトル歪曲係数が 3 個の中心となったならば、S 1 7 において最適歪曲係数を 3 個の中心値に決定して処理を終了する。

【 0 0 3 5 】

音声認識部 6 では、ユーザーの発声を音響分析して得られる L P C ケプストラム係数のうち、第 2 語目の発声の部分を音響分析部 2 から受け取ると同時に、特徴別パターン選択構成部 4 で決定された被選択特徴別パターンと制御対象機器を示す語彙の認識結果及び話者適応処理部 5 で決定されたスペクトル周波数歪曲係数を話者適応処理部 5 から受け取る。そして、ユーザーの発声を音響分析して得られる L P C ケプストラム係数のうち、第 2 語目の発声の部分に対して、話者適応処理部 5 で決定されたスペクトル周波数歪曲係数を用いてスペクトル周波数の歪曲演算を行い、被選択特徴別パターンを用いて制御対象機器を示す語彙の認識結果に応じた認識対象辞書を用いて音声認識処理を行う。

【 0 0 3 6 】

制御信号出力部 7 は、特徴別パターン選択構成部 4 で決定される制御対象機器を示す語彙の認識結果と、音声認識部 6 で決定される制御内容語彙の認識結果を音声認識部 6 から受け取り、制御対象となっている機器に制御内容を示す信号を出力する。

【 0 0 3 7 】

なお、本発明の音声制御装置のうち、サンプリング周波数を 1 0 k H z とし、制御対象機器を「テレビ」の 1 語のみに絞って制御内容語彙を「1 チャンネル」、「音大きく」などの 3 0 単語用意した場合の、S / N 比 1 5 d B の騒音環境下における性能を、音声認識率を用いて（表 1）に示す。

【 0 0 3 8 】

【表 1】

適応方法 \ ユーザーの年齢	12 歳以下	13～64 歳	65 歳以上
適応なし	78.7	88.7	82.3
パターン選択のみによる適応	84.4	89.2	84.5
スペクトル周波数の歪曲のみによる適応	86.8	93.9	83.9
ユーザー音声によるパターン選択と スペクトル周波数の歪曲の併用による適応	90.0	94.6	87.5

【0 0 3 9】

（表 1）の適応方法の欄について説明する。

適応なし、の場合は、学習パターンは単一で、学習対象話者全員を分類せずに使用して作成した場合であって、ユーザー音声のスペクトル周波数の歪曲も行っていない。パターン選択のみによる適応、の場合は、学習パターンを話者の年齢別に作成して、特徴別パターン選択構成部 4 の処理によって、特徴別パターンを選択するが、ユーザー音声のスペクトル周波数の歪曲は行っていない。スペクトル周波数の歪曲のみによる適応、の場合は、学習パターンは適応なしの場合と同様に単一で、学習対象話者全員を分類せずに使用して作成した場合であって、ユーザー音声のスペクトル周波数の歪曲のみを話者適応処理部 5 の処理によって行った場合である。そして、ユーザー音声によるパターン選択とスペクトル周波数の歪曲の併用による適応、の場合には、学習パターンを話者の年齢別に作成して、特徴別パターン選択構成部 4 の処理によって、特徴別パターンを選択し、かつ、ユーザー音声のスペクトル周波数の歪曲を話者適応処理部 5 の処理によって行った場合である。

【0 0 4 0】

一方、ユーザーの年齢を 12 歳以下、13～64 歳、65 歳以上に分けて、本発明の効果について比較したが、（表 1）の結果によって、どの年齢層のユーザーに対しても、本発明の方法の有効性が明らかであることが示された。

【0 0 4 1】

【発明の効果】

本発明によれば、複数存在する特徴別学習パターンの選択のみや、入力音声の

スペクトル周波数の歪曲のみによる話者適応よりも、より性能が高い認識結果が得られるという有利な効果が得られる。

【図面の簡単な説明】

【図 1】

本発明の実施の形態 1 における音声制御装置のブロック構成図

【図 2】

特徴別パターン格納部 3 の詳細構成図

【図 3】

特徴別パターン選択構成部 4 のフローチャート

【図 4】

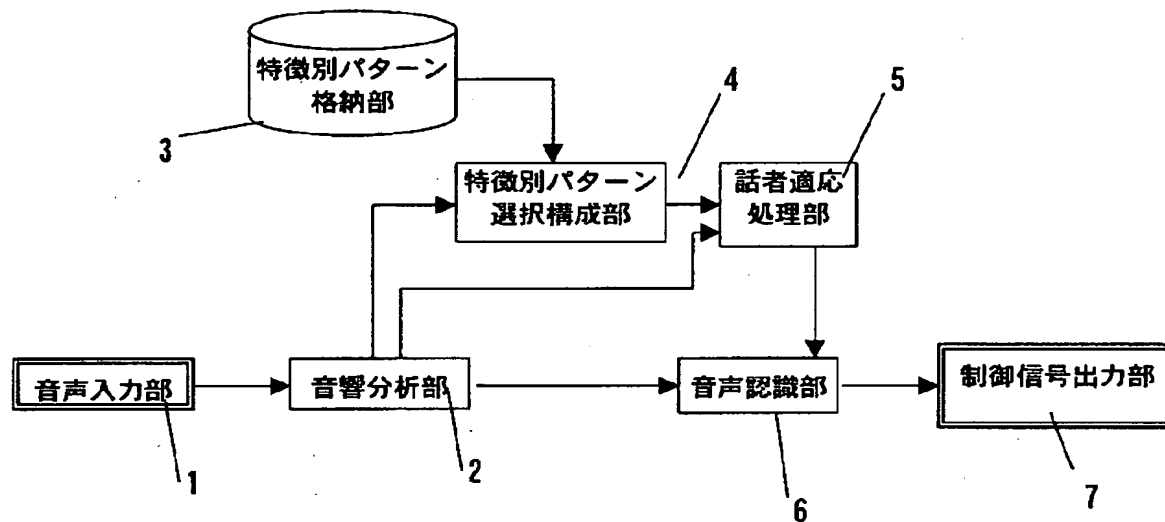
話者適応処理部 5 のフローチャート

【符号の説明】

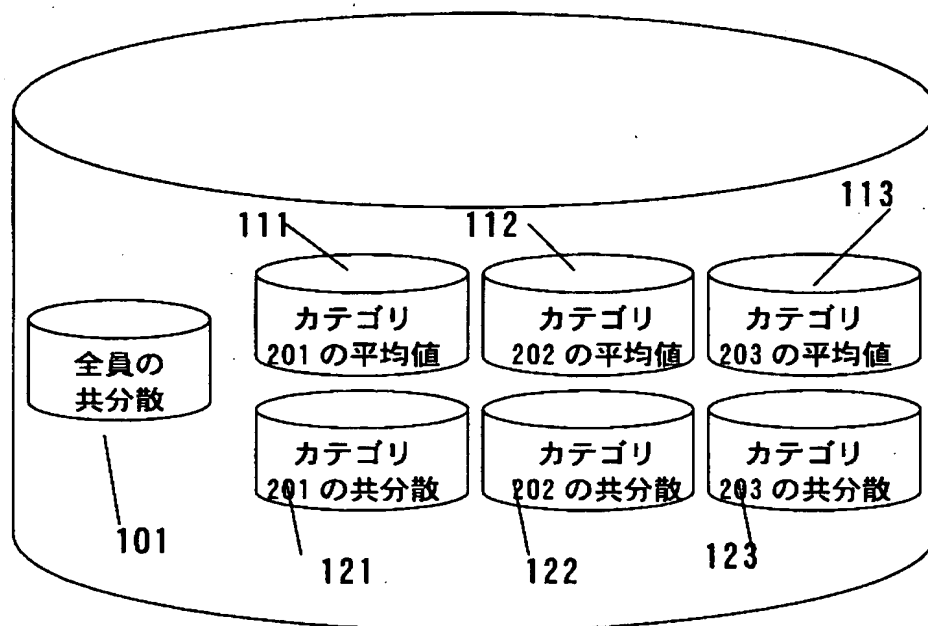
- 1 音声入力部
- 2 音響分析部
- 3 特徴別パターン格納部
- 4 特徴別パターン選択構成部
- 5 話者適応処理部
- 6 音声認識部
- 7 制御信号出力部
- 1 0 1 全員の共分散
- 1 1 1 カテゴリ番号 2 0 1 の平均値
- 1 1 2 カテゴリ番号 2 0 2 の平均値
- 1 1 3 カテゴリ番号 2 0 3 の平均値
- 1 2 1 カテゴリ番号 2 0 1 の共分散
- 1 2 2 カテゴリ番号 2 0 2 の共分散
- 1 2 3 カテゴリ番号 2 0 3 の共分散

【書類名】 図面

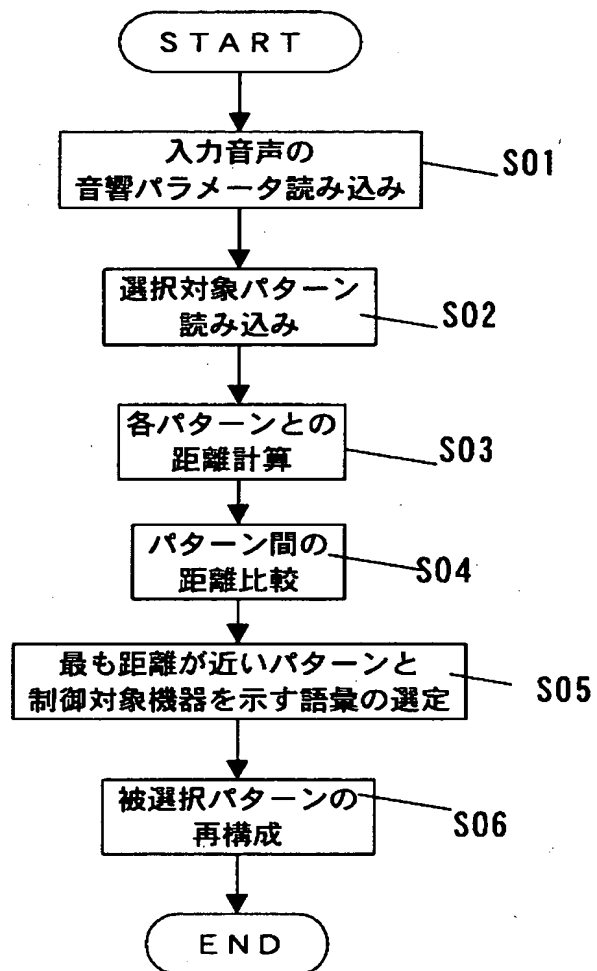
【図 1】



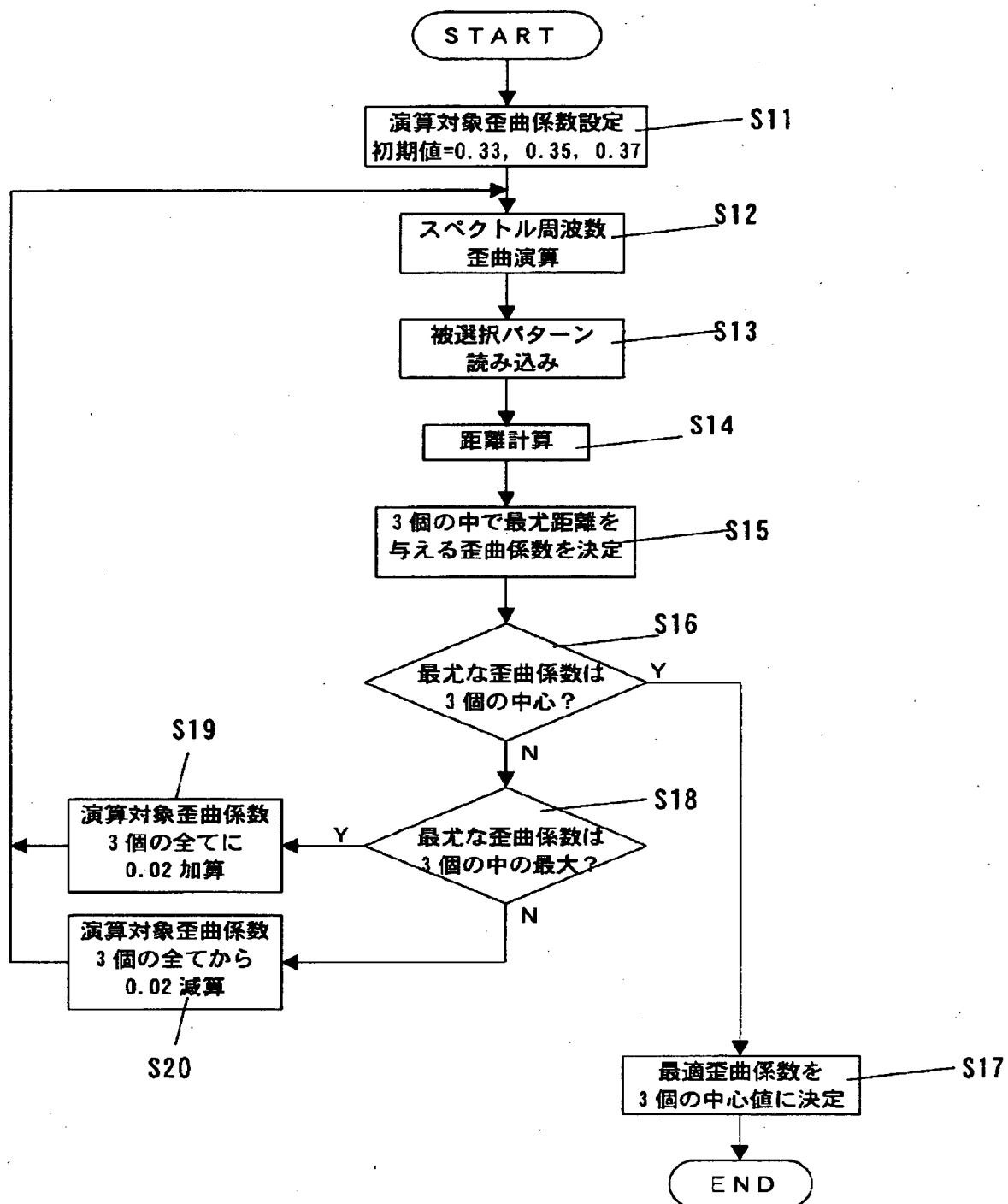
【図 2】



【図3】



【図 4】



【書類名】 要約書

【要約】

【課題】 本発明は、話者の数少ない発声で性能の高い話者適応を行う音声認識方法を提供することを目的とする。

【解決手段】 音声入力部 1 より入力され、音響分析部 2 で音響分析されるユーザーの音声を用いて、あらかじめ学習話者を特徴別に分類して学習した複数の学習パターンを保持する特徴別パターン格納部 3 に格納されている特徴別パターンから、特徴別パターン選択構成部 4 においてユーザーの音声に最適なパターンを選択し、パターン学習話者と入力話者の音声の声道長に起因するスペクトルの周波数の差異を補正するためのスペクトル周波数歪曲係数を話者適応処理部 5 で決定することにより、ユーザーへの小さい負担で後続の音声認識性能を向上させる。

【選択図】 図 1

出 願 人 履 歷 情 報

識別番号 [000005821]

1. 変更年月日 1990年 8月28日
、 [変更理由] 新規登録
住所 大阪府門真市大字門真1006番地
氏名 松下電器産業株式会社